The Harvard Clean Energy Project: Automated, high-throughput screening of organic photovoltaic materials using first-principles electronic structure theory

Johannes Hachmann, Roberto Olivares-Amaya, Sule Atahan-Evrenk, Alán Aspuru-Guzik
Department of Chemistry and Chemical Biology, Harvard University
Cambridge, MA 02138, USA

We present the Harvard Clean Energy Project (CEP, http://cleanenergy.harvard.edu) which is concerned with the computational screening and design of new materials for carbon-based photovoltaics. Plastic solar cells are one of the promising approaches to establishing a ubiquitous source of renewable energy and, e.g., to providing basic electricity to rural areas in undeveloped countries. The necessary 10-15% energy conversion efficiency and lifetime to make this technology commercially viable has, however, not been achieved yet.

With CEP we have established an automated, high-throughput, *in silico* framework to study millions of potential material candidates. Our current project phase is focused on the relevant molecular properties of these candidates – for small molecule setups or as polymer building blocks. We employ first-principles computational quantum chemistry at an unprecedented scale to characterize a wide range of molecular motifs and assess their quality with respect to applications as electronic materials. We complement this high-level effort with techniques from cheminformatics, pattern recognition, and machine learning. In addition to finding specific structures with certain properties, it is the goal of CEP to illuminate and understand the structure-property relations in the domain of organic electronics. Such insights can open the door to a rational, systematic, and accelerated development of future high-performance materials.

CEP is presumably the largest quantum chemical study of all time and it utilizes the massive computational resource of IBM's World Community Grid (www.worldcommunitygrid.org). In collaboration with IBM we have ported the Q-Chem program package to the BOINC environment for distributed, grid-based use on volunteer hosts. In this context, it is deployed as a screensaver application to harvest idle computing time on donor machines, similar to the seti@home, qmc@home, or folding@home projects. This cyberinfrastructure paradigm has already allowed us to characterize 2.2 million molecular graphs of interest in 140 million DFT calculations with a data volume of 300TB. The results are compiled and analyzed in an extensive reference database which will be made available for public use in the spring of 2013. Our presentation addresses data mining, analysis, and scoring, as well as promising molecular motifs that have emerged from CEP so far.

- J. Hachmann, R. Olivares-Amaya, S. Atahan-Evrenk, C. Amador-Bedolla, R.S. Sánchez-Carrera, A. Gold-Parker, L. Vogt, A.M. Brockway, A. Aspuru-Guzik, *J. Phys. Chem. Lett.* 2 (**2011**), 2241.
- R. Olivares-Amaya, C. Amador-Bedolla, J. Hachmann, S. Atahan-Evrenk, R.S. Sánchez-Carrera, L. Vogt, A. Aspuru-Guzik, *Energy Environ. Sci.* 4 (**2011**), 4849.